Automatic classification of gait in children with early-onset ataxia or developmental coordination disorder and controls using inertial sensors

Mannini, Andrea; Martinez-Manzanera, Octavio; Lawerman, Tjitske F.; Trojaniello, Diana; Della Croce, Ugo; Sival, Deborah A.; Maurits, Natasha M.; Sabatini, Angelo Maria

Published in:
Gait & Posture

DOI:
10.1016/j.gaitpost.2016.12.002

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2017

Abstract

Early-Onset Ataxia (EOA) and Developmental Coordination Disorder (DCD) are two conditions that affect coordination in children. Phenotypic identification of impaired coordination plays an important role in their diagnosis. Gait is one of the tests included in rating scales that can be used to assess motor coordination.

A practical problem is that the resemblance between EOA and DCD symptoms can hamper their diagnosis. In this study we employed inertial sensors and a supervised classifier to obtain an automatic classification of the condition of participants. Data from shank and waist mounted inertial measurement units were used to extract features during gait in children diagnosed with EOA or DCD and age-matched controls. We defined a set of features from the recorded signals and we obtained the optimal features for classification using a backward sequential approach. We correctly classified 80.0%, 85.7%, and 70.0% of the control, DCD and EOA children, respectively. Overall, the automatic classifier correctly classified 78.4% of the participants, which is slightly better than the phenotypic assessment of gait by two pediatric neurologists (73.0%). These results demonstrate that automatic classification employing signals from inertial sensors obtained during gait maybe used as a support tool in the differential diagnosis of EOA and DCD. Furthermore, future extension of the classifier's test domains may help to further improve the diagnostic accuracy of pediatric coordination impairment. In this sense, this study may provide a first step towards incorporating a clinically objective and viable biomarker for identification of EOA and DCD.

Keywords

Gait; Early-Onset Ataxia; Developmental Coordination Disorder; Inertial Sensors; Accelerometers; Gyroscopes

Introduction

Coordination is characterized by smooth and efficient goal directed movements that involve different parts of the body. Correct anticipation and knowledge of where the body is located in space (proprioception) are essential for the execution of motor tasks requiring coordination. The cerebellum plays a pivotal role in the organization of planned coordination. It integrates input from different motor and multisensory feedback signals of different body regions. Gait requires a complex interaction of different muscles to maintain balance while moving forward, and even though children start to walk around their first year of age, it continues developing at least until the age of eleven [1]. Gait therefore can be affected by impaired coordination. In children, coordination can be affected due to different causes, such as ataxia, developmental coordination disorder (DCD) and physiological immaturity of the cerebellar circuitry in young children. Early-Onset Ataxia (EOA) is characterized by chronically impaired coordination of voluntary, goal directed movements starting before the 25th year of life[2–5]. The underlying etiology is either associated with dysfunctional cerebellar networks or with abnormal spinal afferent input. Many of the heterogeneous underlying genetic causes of EOA will show progression over time, resulting in wheelchair dependency and even shorter life expectancy [6]. DCD is characterized by abnormal coordination impairment, after the exclusion of medical (behavioral or neurological) conditions as the underlying cause. DCD may involve impaired acquisition of motor skills, sensorimotor integration, postural control, strategic planning, visual-spatial processing and executive functioning[7–9]. Although the future perspective of DCD is much more optimistic compared to EOA, patients diagnosed with DCD may experience motor difficulties even into adulthood [7]. With treatment, functional outcome in these children can be improved[7]. Finally, in young healthy children (CTRL), immaturity of the cerebellar circuitry is characterized by normal, physiologically immature coordination, with features that can mimic “ataxia” [5,10]. As implicated by the descriptions, these three clinical entities for

coordination impairment are characterized by overlapping features, which potentially hampers unanimous phenotypic recognition. However, due to the different future perspectives and treatment options, early distinction between EOA, DCD and healthy controls is desirable. In addition, adequate distinction between EOA and DCD will hopefully improve the yield of innovative genetic strategies and enhance the quality of data entry in international EOA databases.

In absence of reliable distinctive biomarkers, the Scale for Assessment and Rating of Ataxia (SARA), is often used as an additional, supportive biomarker to indicate ataxia severity [11]. Despite the high reliability of the scale, we have shown that pediatric SARA is confounded by other factors than ataxia, as well [5,10,12]. Nevertheless, we have shown that the relative SARA gait subscore can support the recognition of an indisputable EOA phenotype in mildly affected participants[13]. Presently available quantitative gait parameters [14] are still not ubiquitously implemented as a clinical tool. Based on the remarks reported above, we reasoned that clinically simple and reproducible quantitative gait analysis could be worthwhile for reliable EOA and DCD recognition.

In the present paper, we evaluate a method for the automatic and objective assessment of pediatric gait as compared to a clinical diagnosis in a similar way to what was previously done for other pathological conditions [15,16]. Additionally, the accuracy of phenotypic assessments is estimated. Both methods classify patients into three groups (EOA, DCD and CTRL). To be effective, the automatic assessment is expected to guarantee both a limited increase of the complexity of the evaluation and a minimal impact on the gait patterns under evaluation. As a result, we chose to apply a supervised classification algorithm to gait kinematics patterns recorded with few, light weight, wearable inertial measurement units (IMUs). Some of the features employed were obtained by modeling gait sequences with

Hidden Markov Models (HMMs), which were shown to be effective in analyzing gait sequence data acquired with IMUs [17–22].

Materials and methods

Participants

The study was performed in accordance with the research and integrity codes of the UMCG. Since gait assessment is routinely performed as part of scoring of the SARA during clinical assessment, the Medical Ethical Committee of the UMCG provided a waiver for ethical approval. After informed consent by the parents and informed assent by the participants (when older than 12 years of age), we included ten EOA [m 13.3 (sd 3.8) years], seven DCD [m 9.6 (sd 2.2) years] and twenty age-matched CTRL [m 12.1 (sd 3.3) years] children. There were no significant age differences between groups (ANOVA, p=0.07).

The inclusion criterion for EOA was clinically assessed ataxia before the 25th year of life, either confirmed by a prior diagnosis and/or confirmed by two specialists from the movement disorders team (with access to the clinical radiologic evaluations, metabolic tests and genetic data). Identified EOA diagnoses involved: Niemann Pick Type C (n=1), MHBD-deficiency (n=1), Friedreich's Ataxia (n=2), CACNA1A (n=2) and unknown (n=4). The inclusion criterion for DCD was the assessment of impaired coordination as clinically established by an independent rehabilitation clinician, according to DSM-IV-TR[23], after exclusion of a movement disorder by a neurologist. The inclusion criteria for healthy young children were the ability to follow mainstream education and absence of any neurological or orthopedic disorder as well as other physical conditions or prescribed medication that could theoretically interfere with the execution of SARA tasks.

The SARA scale represents an ataxia rating scale in the domains of gait, upper limbs coordination, and speech, with scores varying from zero (no ataxia) to the maximum severity

of 40[11]. The SARA gait subscore varies from zero (no difficulties in walking) to eight (unable to walk). We compared SARA score and SARA gait subscore between groups using an ANOVA test in case of normally distributed data and a Kruskal-Wallis test for non-normally distributed data.

During their visit to the UMCG outpatient clinic, we videotaped the SARA performances of all participants. The SARA gait evaluation consists of the assessment of 1) walking at a safe distance parallel to a wall and 2) walking in tandem without support [11]. In this study we focused on 1). According to SARA guidelines [11], participants were asked to walk in a straight line at their own speed in a corridor of approximately 15 meters, turn 180^{o} and return to the starting position. We strived to obtain a similar number of strides and trials from all participants. However, due to their condition, the number of recorded strides varied across participants. In particular, the gait segmentation algorithm identified 54.4±17.3 strides (mean ±standard deviation) for control subjects, 53.6±12.8 strides for DCD patients and 40.9±16.9 strides for EOA patients. These performances were recorded by six IMUs (Shimmer3, Shimmer, Dublin, Ireland) including three accelerometers and three gyroscopes that were attached to the body with elastic straps. Data were recorded at a sampling rate of 256 Hz while participants performed the tasks described in the SARA. Before each recording IMUs were calibrated using software from the manufacturer (Shimmer 9DoF Calibration v2.5). One IMU was placed on the sternum, another one on the low back close to the L3 vertebra, two were placed bilaterally halfway each upper leg over the quadriceps and two on the lateral side of the shanks, just above the malleolus. This set-up was chosen to be able to carry out various analyses including joint kinematics analysis during SARA motor tasks. However, given the goal of this study, only data from a subset of IMUs was used.

Clinical diagnosis

Three experienced pediatric assessors (two pediatric neurologists and a movement disorders investigator specialized in ataxia) performed quantitative SARA assessments. Previous publications have shown that the SARA score is reliable when assessed by this group [5,12].

Phenotypic assessment

After a time interval of six months, two pediatric neurologists independently assessed the phenotypic characteristics of the videotaped SARA gait performances and assigned the children to the EOA, DCD and CTRL gait-subgroups. Prior to assessment, the pediatric neurologists did not have access to the clinical or previous scoring data.

Automatic classification

The automatic classification of patients into the three groups (CTRL, DCD and EOA) was carried out similarly to a previous work [17], in which data from three groups (healthy elderly, hemiparetic patients and patients with Huntington's disease) were classified based on data obtained from wearable IMUs. Seven IMU-derived signals were selected among the IMU-derived signals recorded during each gait trial. Six of them were extracted from an IMU positioned on the shank: the medio-lateral (ML) angular velocity and its approximated derivative, the antero-posterior (AP) acceleration and its approximated derivative, the approximated derivatives of the ML and vertical (VT) accelerations[14], [20]. A single signal was extracted from the IMU attached to the lower back (the ML acceleration) [17]. The above mentioned signals were selected since, in a previous work, they were found to be suitable for recognizing gait alterations, [17]. A schematic of the method is presented in figure 1.

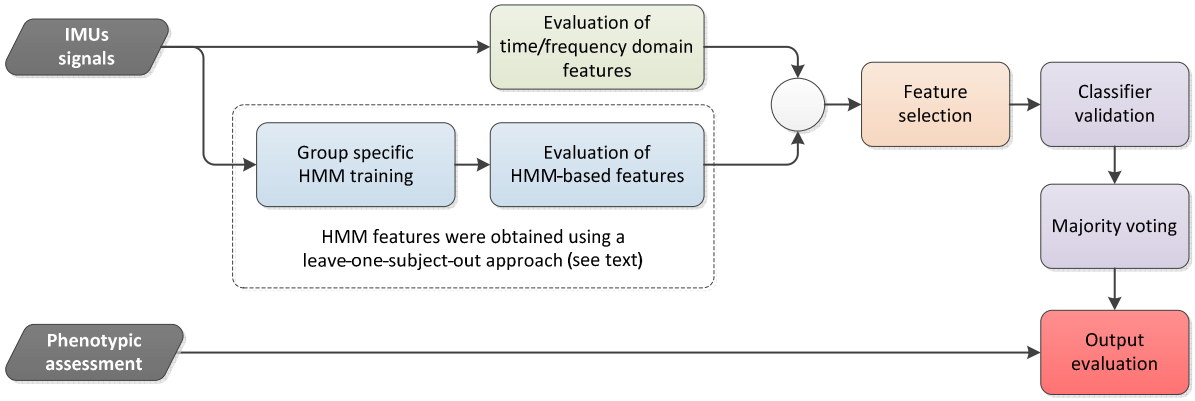

Figure 1: schematic summarizing the proposed methodology for automatic classification. Variables containing information suitable for classification purposes (features) were extracted from the selected IMU signals. An automatic procedure to reduce the complexity of the classification problem was included (feature selection) and then an automatic classifier was validated. The results of each tested walking trial were then summarized to provide a single output for each subject (with the majority voting method). The results were then compared to the phenotypic assessment.

The signals derived from the shank-mounted IMUs recorded during the walking trials were processed to extract classification features. From each of the selected IMU signals six features in the time domain and six in the frequency domain were extracted (12x7 = 84 features). Six additional features were obtained by modeling gait sequences using Hidden Markov Models (HMMs) [17](Table 1). HMMs are a pattern recognition method that provides a statistical framework for modeling signals [25]: the resulting signal models can be specific to particular conditions and can then be used to classify new data by evaluating which specific model better explains new data (model *likelihood* evaluation). HMMs were trained in a supervised way by pairing stance and swing phases of gait to model *states*[22]. Reference gait events were extracted from the IMU signals using a previously validated method for gait segmentation [26]. To obtain HMM-based features, a model was trained for each of the three groups at each validation step. In particular, the data likelihood under each model was evaluated and provided six additional features to be used for classifying data during the testing phase: three features were obtained by the evaluation of model likelihoods

on 2-second windows of data and three features were obtained by comparing the likelihood evaluated across the full length of the walking trials[17].

Features that did not improve the classification accuracy were sequentially discarded, one at a time, by means of an automatic method, the sequential backward feature selection [27]. The cross-validation accuracy was used as the criterion for each selection step.

After the automatic selection of a subset of the original 90 features, a classification algorithm was applied. A support vector machine (SVM) classifier with a radial basis function kernel was used. Classifier parameters were retained from a previous work: the upper complexity bound and the kernel variability were fixed to C = 100 and γ= 0.01, respectively [17]. The classifier was trained using a weighted cost function to limit the effect of class unbalance using the *LibSVM* implementation.[28]

A leave-one-subject-out (LOSO) cross-validation was performed for both training phases (HMMs and SVM). At each validation step, data from one participant were excluded from the training set and the solution obtained was tested on data from the excluded participant. This was then repeated to test all participants in the dataset and results were aggregated by summing the confusion matrices obtained at each step.

The results obtained from the SVM classifier referred to single walking trials. To classify a patient, a majority voting strategy [27] was applied for which the classification output of each side in each walking trial generated a vote. The class collecting most votes was then selected as the winner of the poll. A heuristic rule to deal with ties was introduced.

Classification accuracy

The accuracy of the phenotypic assessment and of the automatic classification are presented as confusion matrices using the clinical diagnosis as reference. The accuracy of the automatic classification was determined for single gait trials and after the majority voting. To facilitate

a comparison between phenotypic assessment and the automatic classification the assessments of both evaluators were aggregated in one confusion matrix.

Results

Participant characteristics

According to Shapiro-Wilk tests, the total SARA score and the gait score were normally distributed in the EOA and DCD groups but not in the CTRL group. Both total SARA and SARA-gait scores differed significantly between groups (Kruskal-Wallis test,$p<0.01$). Post-hoc Mann-Whitney U tests revealed that SARA total scores were significantly higher in EOA than in DCD ($p<0.01$) and that SARA total scores were significantly higher in DCD than in CTRL ($p<0.01$). SARA gait scores were significantly different between groups, as well (Kruskal-Wallis test,$p<0.01$). Post-hoc Mann-Whitney U tests showed that SARA gait scores were significantly higher in EOA than in DCD ($p<0.01$) and that SARA gait scores were significantly higher in DCD than in CTRL ($p<0.05$) (Table 2).

Feature selection

The feature set obtained by applying the sequential backward feature selection is summarized in Table 1. Retained features are indicated with check marks: four out of six HMM-based features and 37 out of 84 features in the time and frequency domain were retained.

Phenotypic assessment and automatic classification results

The confusion matrix for the phenotypic assessment performed by the two specialists is reported in Table 3, part 1. Every assessment is reported as an entry for the confusion matrix. The SVM classifier output is summarized in confusion matrices reported in Table 3 (parts 2 and 3). The first classification output describes the walking trials classification and shows that 63.8% of walking trials were assigned to the correct group. The majority voting resulted

in correct classification for 78.4% of participants. In particular, no DCD or EOA participants were incorrectly classified as CTRL. However, four CTRL participants were mistakenly classified as DCD or EOA and a few misclassifications occurred between EOA and DCD.

The relation between the SARA gait subscore and the classifier output is shown in figure 2.

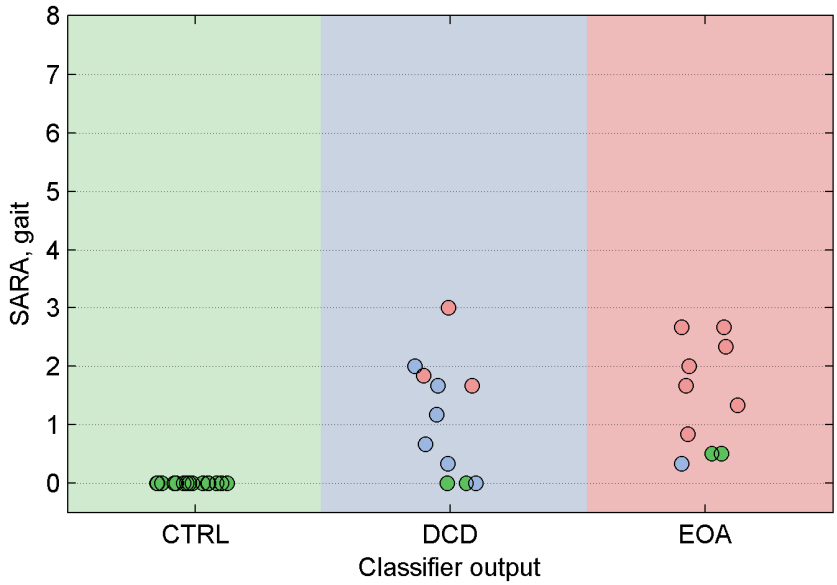

Figure 2: SARA gait scores for controls (green dots), DCD (blue dots) and EOA (red dots) participants in relation to the output of the classifier.

Discussion

In this study, we presented an automatic assessment of gait in EOA, DCD and healthy "immature" control children that could support the currently challenging phenotypic recognition of such conditions. We implemented a method for the automatic classification of wearable inertial sensors data from gait trials according to three categories of coordination impairment. To assess the accuracy of the method, we compared the quantitative gait outcomes with: (a) the clinical diagnosis based on genetic, radiologic, neurologic and/or metabolic data; and (b) the observed phenotype (as determined by clinical specialists of movement disorders).To assess the accuracy of the method, we determined the percentage of correct classification and of misclassification using the clinical diagnosis as reference. We then compared the accuracy of the automatic classification method with that of the

phenotypic assessment as determined by clinical specialists in movement disorders. We reasoned that if quantitative gait features are able to distinguish between EOA, DCD and CTRL groups, this technique could provide an objective tool for the identification of EOA and DCD. Overall, the classifier obtained an accuracy of 78.4%, which is 5.4% higher than the mean accuracy of the phenotypic assessment. From these data, we conclude that the quantitative gait features, as provided by the automatic classifier, can provide a supportive tool for unanimous and reproducible diagnostic assessment.

For the purpose of this discussion we looked into some individual misclassification cases. The automatic classifier placed one DCD participant in the EOA group. This is most probably due to the accidental misplacement of the shank mounted IMUs, occurred in the participant data acquisition session and recognized by analyzing the video-recordings of his walking trials. Interestingly, this participant was placed in the CTRL group by both evaluators. The automatic classifier misclassified three EOA and two CTRL participants, placing them in the DCD group. In two of these cases, one of the two evaluators agreed with the automatic classifier. There were two cases of misclassified CTRL participants in the EOA group. These participants were placed in the EOA and DCD groups and in the DCD and CTRL groups respectively by the evaluators. They also obtained impaired/sub-optimal SARA gait subscores, suggesting that phenotypical assessment and automatic classification identified a sub-optimal/impaired coordination.

Comparing the automatic classifier assessment with the phenotypic assessment revealed a higher diagnostic accuracy by the former in DCD subjects (50% higher) and a higher diagnostic accuracy by the latter in EOA patients (20% higher). For controls, both methods revealed similar accuracies, with a slightly higher accuracy of the automatic classifier (2.5 % higher).

To explain the outcomes of this study, it is crucial to elaborate on the characteristics of the three methods of classifications utilized: the clinical diagnosis, the phenotypic assessment, the automatic method consideration. The clinical diagnosis is the result of the evaluation of all potentially useful parameters. This may implicate that indicators other than gait parameters (such as genetic, radiologic, laboratory) could have been decisive for the clinical diagnosis. From this perspective, a different classification between the clinical diagnosis and the automatic gait classifier does not necessarily imply a poor performance of the classifier. For instance, a child with a genetic diagnosis and discrete changes in tandem gait does not necessarily reveal abnormalities in the walking pattern that can be picked up by the classifier. Similarly, the phenotypic assessment which is based on videotaped SARA gait performances could be heavily affected by the observation of tandem gait, standing and by the perception of the age of the child, expressions that were not included in the recordings processed by the automatic classification, which is applied only to data recorded during straight walking. Considering that the automatic classifier was applied exclusively to straight gait recordings, a 78% classification accuracy is very promising. Once the automatic classifier application will be extended to other SARA gait and kinetic parameters, it is expected that the accuracy of this method will increase.

Interestingly, the phenotypic assessment revealed a higher sensitivity for EOA patients, whereas the automatic classifier revealed a higher sensitivity for DCD and control subjects. As EOA represents a neurologic diagnosis, and as DCD represents a practical rehabilitation diagnosis (after exclusion of neurologic abnormalities), it appears hardly surprising that pediatric neurologists are better skilled to identify EOA than DCD. As specific standards for DCD recognition are still missing, it appears tempting to speculate that future classifier-based assessments of additional DCD domains may assist further delineation of this broad diagnostic group. Within the limitation of the present study, we would thus suggest that

future extension of the classifier's test domains and also inclusion of a larger number of patients may help to improve the diagnostic accuracy of pediatric coordination impairment. Hopefully, this study provides a first step towards incorporating a clinically objective and viable biomarker for uniform identification of EOA and DCD.

Conflicts of Interest

The authors of this manuscript certify that there is no conflict of interest with any financial or non-financial organization or entity regarding the material discussed in the manuscript.

References

[1] S.J. Hillman, B.W. Stansfield, A.M. Richardson, J.E. Robb, Development of temporal and distance parameters of gait in normal children., Gait Posture. 29 (2009) 81–5. doi:10.1016/j.gaitpost.2008.06.012.

[2] M. Mumenthaler, H. Mattle, Fundamentals of Neurology An Illustrated Guide, 1st ed., Thieme Medical Publishers, 2006.

[3] A.E. Harding, Classification of the hereditary ataxias and paraplegias., Lancet (London, England). 1 (1983) 1151–5. http://www.ncbi.nlm.nih.gov/pubmed/6133167 (accessed December 6, 2015).

[4] A. Chio, L. Orsi, P. Mortara, D. Schiffer, Early onset cerebellar ataxia with retained tendon reflexes: prevalence and gene frequency in an Italian population., Clin. Genet. 43 (1993) 207–11. http://www.ncbi.nlm.nih.gov/pubmed/8330454 (accessed January 28, 2016).

[5] R. Brandsma, A.H. Spits, M.J. Kuiper, R.J. Lunsing, H. Burger, H.P. Kremer, D.A.

316 Sival, Ataxia rating scales are age-dependent in healthy children., Dev. Med. Child
317 Neurol. 56 (2014) 556–63. doi:10.1111/dmcn.12369.

318 [6] B.L. Fogel, S. Perlman, Clinical features and molecular genetics of autosomal
319 recessive cerebellar ataxias, Lancet Neurol. 6 (2007) 245–257.

320 [7] J.G. Zwicker, C. Missiuna, S.R. Harris, L.A. Boyd, Developmental coordination
321 disorder: a review and update., Eur. J. Paediatr. Neurol. 16 (2012) 573–81.
322 doi:10.1016/j.ejpn.2012.05.005.

323 [8] J.G. Zwicker, C. Missiuna, L.A. Boyd, Neural correlates of developmental
324 coordination disorder: a review of hypotheses., J. Child Neurol. 24 (2009) 1273–81.
325 doi:10.1177/0883073809333537.

326 [9] P.H. Wilson, S. Ruddock, B. Smits-Engelsman, H. Polatajko, R. Blank, Understanding
327 performance deficits in developmental coordination disorder: a meta-analysis of recent
328 research., Dev. Med. Child Neurol. 55 (2013) 217–28. doi:10.1111/j.1469-
329 8749.2012.04436.x.

330 [10] D.A. Sival, E.R. Brunt, The International Cooperative Ataxia Rating Scale shows
331 strong age-dependency in children., Dev. Med. Child Neurol. 51 (2009) 571–2.
332 doi:10.1111/j.1469-8749.2009.03334.x.

333 [11] T. Schmitz-Hübsch, S. Du Montcel, S.T. du Montcel, L. Baliko, J. Berciano, S.
334 Boesch, C. Depondt, P. Giunti, C. Globas, J. Infante, J.-S. Kang, B. Kremer, C.
335 Mariotti, B. Melegh, M. Pandolfo, M. Rakowicz, P. Ribai, R. Rola, L. Schöls, S.
336 Szymanski, B.P. van de Warrenburg, A. Dürr, T. Klockgether, R. Fancellu, Scale for
337 the assessment and rating of ataxia Development of a new clinical scale, Neurology.
338 66 (2006) 1717–20. doi:10.1212/01.wnl.0000219042.60538.92.

339 [12] D.A. Sival, M.E. Pouwels, A. Van Brederode, N.M. Maurits, C.C. Verschuuren-

340 Bemelmans, E.R. Brunt, G.J. Du Marchie Sarvaas, R.J. Verbeek, O.F. Brouwer, J.H.
341 Van Der Hoeven, In children with Friedreich ataxia, muscle and ataxia parameters are
342 associated., Dev. Med. Child Neurol. 53 (2011) 529–34. doi:10.1111/j.1469-
343 8749.2011.03931.x.

344 [13] T.F. Lawerman, R. Brandsma, J.T. van Geffen, R.J. Lunsing, H. Burger, M.A.J.
345 Tijssen, J.J. de Vries, T.J. de Koning, D.A. Sival, Reliability of phenotypic early-onset
346 ataxia assessment: a pilot study., Dev. Med. Child Neurol. (2015) 1–7.
347 doi:10.1111/dmcn.12804.

348 [14] A. Muro-de-la-Herran, B. Garcia-Zapirain, A. Mendez-Zorrilla, Gait analysis methods:
349 an overview of wearable and non-wearable systems, highlighting clinical applications.,
350 Sensors (Basel). 14 (2014) 3362–94. doi:10.3390/s140203362.

351 [15] G. Pendharkar, P. Percival, D. Morgan, D. Lai, Automated method to distinguish toe
352 walking strides from normal strides in the gait of idiopathic toe walking children from
353 heel accelerometry data, Gait Posture. 35 (2012) 478–482.
354 doi:10.1016/j.gaitpost.2011.11.011.

355 [16] N. O'Mahony, B. Florentino-Liano, J.J. Carballo, E. Baca-García, A.A. Rodríguez,
356 Objective diagnosis of ADHD using IMUs, Med. Eng. Phys. 36 (2014) 922–926.
357 doi:10.1016/j.medengphy.2014.02.023.

358 [17] A. Mannini, D. Trojaniello, A. Cereatti, A.M. Sabatini, A Machine Learning
359 Framework for Gait Classification Using Inertial Sensors: Application to Elderly, Post-
360 Stroke and Huntington's Disease Patients., Sensors (Basel). 16 (2016).
361 doi:10.3390/s16010134.

362 [18] N. Abaid, P. Cappa, E. Palermo, M. Petrarca, M. Porfiri, Gait detection in children
363 with and without hemiplegia using single-axis wearable gyroscopes., PLoS One. 8

(2013) e73152. doi:10.1371/journal.pone.0073152.

[19] J. Taborri, S. Rossi, E. Palermo, F. Patanè, P. Cappa, A novel HMM distributed classifier for the detection of gait phases by means of a wearable inertial sensor network., Sensors (Basel). 14 (2014) 16212–34. doi:10.3390/s140916212.

[20] M. Chen, B. Huang, Y. Xu, Human Abnormal Gait Modeling via Hidden Markov Model, 2007 Int. Conf. Inf. Acquis. (2007) 517–522. doi:10.1109/ICIA.2007.4295787.

[21] A. Mannini, A.M. Sabatini, Walking speed estimation using foot-mounted inertial sensors: Comparing machine learning and strap-down integration methods, Med. Eng. Phys. 36 (2014) 1312–1321. doi:10.1016/j.medengphy.2014.07.022.

[22] A. Mannini, A.M. Sabatini, Gait phase detection and discrimination between walking-jogging activities using hidden Markov models applied to foot motion data from a gyroscope., Gait Posture. 36 (2012) 657–61. doi:10.1016/j.gaitpost.2012.06.017.

[23] American Psychiatry Association, Diagnostic and Statistical Manual of Mental Disorders, 4th ed., American Psychiatric Publishing, Inc, Washington D.C., 2000. doi:10.1176/appi.books.9780890425596.893619.

[24] A. Mannini, D. Trojaniello, U. Della Croce, A.M. Sabatini, Hidden Markov model-based strategy for gait segmentation using inertial sensors: Application to elderly, hemiparetic patients and Huntington's disease patients., Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf. 2015 (2015) 5179–82. doi:10.1109/EMBC.2015.7319558.

[25] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE. 77 (1989) 257–286. doi:10.1109/5.18626.

[26] D. Trojaniello, A. Cereatti, E. Pelosin, L. Avanzino, A. Mirelman, J.M. Hausdorff, U. Della Croce, Estimation of step-by-step spatio-temporal parameters of normal and

impaired gait using shank-mounted magneto-inertial sensors: application to elderly, hemiparetic, parkinsonian and choreic gait., J. Neuroeng. Rehabil. 11 (2014) 152. doi:10.1186/1743-0003-11-152.

[27] A.K. Jain, R.P.W. Duin, J. Mao, Statistical Pattern Recognition : A Review, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 4–37.

[28] C. Chang, C. Lin, LIBSVM : A Library for Support Vector Machines, ACM Trans. Intell. Syst. Technol. 2 (2013) 1–39. doi:10.1145/1961189.1961199.

Table 1. Feature set for classification, adapted from[17]. Check marks indicate the feature set after the automatic feature selection strategy. Data channels obtained by approximated derivatives are indicated with a *"d"*.

Category	**Feature**	**Channel** *(a check mark indicates that the feature has been selected by the automatic selection strategy)*						
		ML gyro	*dML gyro*	*AP acc.*	*dAP acc.*	*dVT acc.*	*dML acc.*	*dML acc. (waist)*
Time Domain Features	Mean value			✓	✓	✓	✓	✓
	Standard deviation			✓		✓		✓
	Variance			✓	✓			
	Maximum		✓		✓	✓		
	Minimum	✓						
	Range					✓	✓	
Frequency Domain Features	Power at first dominant freq. (P1)	✓			✓			
	Power at second dominant freq.		✓			✓	✓	
	First dominant frequency				✓	✓	✓	✓
	Second dominant frequency			✓	✓	✓	✓	✓
	Total power (PT)		✓				✓	
	P1 / PT			✓	✓	✓	✓	✓
HMM-Based Features *(Models use data from all the seven data channels)*	Log-likelihood, CTRL model (limited to a 2-s window)				✓			
	Log-likelihood, DCD model (limited to a 2-s window)							
	Log-likelihood, EOA model (limited to a 2-s window)							
	Difference between log-likelihoods for CTRL and DCD models (for all available data)				✓			
	Difference between log-likelihoods for CTRL and EOAmodels(for all available data)				✓			
	Difference between log-likelihoods for DCD and EOAmodels(for all available data)				✓			

Table 2. Participant characteristics.

		CTRL (20)	DCD (7)	EOA (10)
Age	**Mean (sd)** **Range**	12.1 (3.3) 7-20	9.6 (2.2) 7-13	13.3 (3.8) 8-19
SARA score	**Median (IQR)** **Range**	0.3 (0.7) 0-2.25	2.5 (4.0) 0.5-11.25	9.1 (6.1) 4.5-17
SARA gait subscore	**Median (IQR)** **Range**	0.0 (0.0) 0-0.5	1.0 (2.0) 0-4	3.5 (2.4) 1-6

Table 3. Confusion matrices for the group classification.To facilitate a comparison between phenotypic assessment and automatic classification the assessments of both evaluators were aggregated. Results obtained by phenotypic assessment are in part 1. Results obtained using the automatic classifier for single walking trials (part 2) and after majority voting (part 3).

		Phenotype classification output					
		CTRL		**DCD**		**EOA**	
1. phenotypic assessment output							
Clinical Diagnosis	**CTRL**	31	(77.5%)	8	(20.0%)	1	(2.5%)
	DCD	4	(28.6%)	5	(35.7%)	5	(35.7%)
	EOA	0 (0.0%)		2	(10.0%)	18	(90.0%)
		Overall accuracy 73.0 % of assessments					

		Automatic classification output					
		CTRL		**DCD**		**EOA**	
2. automatic classification output for single walking trials							
Clinical Diagnosis	**CTRL**	107	(61.5%)	39	(22.4%)	28	(16.1%)
	DCD	5	(8.3%)	36	(60.0%)	19	(31.7%)
	EOA	1	(1.4%)	18	(25.7%)	51	(72.9%)
		Overall accuracy 63.8% of walking episodes					
3. automatic classification output after majority voting							
Clinical Diagnosis	**CTRL**	16	(80.0%)	2	(10.0%)	2	(10.0%)
	DCD	0	(0%)	6	(85.7%)	1	(14.3%)
	EOA	0	(0%)	3	(30.0%)	7	(70%)
		Overall accuracy 78.4% of participants					